

Linkage Disequilibrium Mapping in Trisomic Populations: Analytical Approaches and an Application to Congenital Heart Defects in Down Syndrome

Kimberly F. Kerstann,¹ Eleanor Feingold,² Sallie B. Freeman,¹ Lora J. H. Bean,¹ Robert Pyatt,¹ Stuart Tinker,¹ Amy H. Jewel,³ George Capone,³ and Stephanie L. Sherman^{1*}

¹Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia

²Departments of Human Genetics and Biostatistics, University of Pittsburgh, Graduate School of Public Health, Pittsburgh, Pennsylvania

³Division of Neurology and Developmental Medicine, Kennedy Krieger Institute, Baltimore, Maryland

Many of the birth defects associated with trisomy exhibit both variable expressivity and incomplete penetrance. This variability suggests that it is allelic variation and not simply the presence of an additional chromosome that leads to the development of certain trisomy-associated birth defects. With the proper tools, one may use trisomic populations to identify genes involved in the development of specific birth defects. A trisomic population may be advantageous over a normal population if the defect is over-represented in the trisomic population. Alternatively, one can view the trisomic populations as a "model system" to offer insight into aspects of both normal and abnormal embryonic development. Standard disomic linkage disequilibrium mapping approaches need to be adjusted to account for the presence of the additional genetic material in the trisomic individuals. We present an approach for linkage disequilibrium mapping of variable phenotypes in a trisomic population that adequately accounts for the additional alleles and the pattern of non-independent inheritance. We establish the laboratory methods and statistical tools necessary to conduct an association study in a trisomic population. As an example, we have applied these tools to a pilot study of Down syndrome-associated congenital heart defects. © 2004 Wiley-Liss, Inc.

Key words: trisomy; association studies; Down syndrome; congenital heart defects

Contract grant sponsor: NIH; Contract grant numbers: P01 HD24605, R01 HD38979; Contract grant sponsor: CRC US DHS NIH; Contract grant number: MO-1-RR-00039.

*Correspondence to: Stephanie L. Sherman, Dept. of Human Genetics, Emory University School of Medicine, 615 Michael St., Whitehead Bldg, Suite 301, Atlanta, GA 30322. E-mail: ssherman@genetics.emory.edu

Received 30 December 2004; Accepted 18 May 2004

Published online 3 August 2004 in Wiley InterScience (www.interscience.wiley.com)

DOI: 10.1002/gepi.20019

INTRODUCTION

Trisomy 21, the chromosomal anomaly responsible for over 95% of Down syndrome (DS), is the most common identified cause of mental retardation in humans with an incidence of approximately 1 in 600 live births [Fryns, 1987]. Although mental retardation is the most recognized phenotype of DS, several associated traits exhibit both variable expressivity and incomplete penetrance. For example, 44% of all individuals with DS have some form of congenital heart defect (CHD) [Freeman et al., 1998], while 8% exhibit some form of gut defect [Epstein, 2001] and 1% develop transient leukemia [Zipursky et al., 1992]. The frequency of these defects among individuals with DS is dramatically increased compared with the

frequency observed in the general population. For example, only 5–7 in 1,000 non-trisomic newborns are estimated to have some form of CHD [Emanuel, 1970]. Atrioventricular septal defects (AVSD), the most common form of CHD in the DS population (45% of the 40% of cases with CHD and DS), are observed in the general population in only 3–4/10,000 live births [Loffredo et al., 2001]. Thus, there is a 500-fold increased risk for AVSD among newborns with DS compared to those without DS.

The excess of certain traits in the DS population suggests an etiology associated with increased dosage of genes on chromosome 21. Knowledge gained through analysis of defects in the DS population would offer insight into aspects of both normal and abnormal human development

in those with and without trisomy. Utilization of a DS population in the analysis of birth defects is also advantageous in that chromosome 21 may be used as a candidate chromosome in the initial scan for genes involved in the defect(s) in question. However, variation in genes on other chromosomes in the context of increased dosage of chromosome 21 should also be considered.

The observation that only a percentage of DS individuals exhibit these variable phenotypes suggests that factors other than general over-expression of genes on chromosome 21 may be involved in the susceptibility of these traits. Three factors have been proposed to explain this variation: (1) stochastic factors, (2) extrinsic factors, and (3) genetics differences [Epstein, 2001]. Most likely, all three mechanisms contribute to phenotypic variation in some fashion. Genetic effects may be the direct result of increased dosage of specific genes on chromosome 21 leading to their altered regulation or gene function. Allelic variants of the chromosome 21 genes themselves or of non-chromosome 21 genes in the trisomic background could be responsible for the development of the variable traits associated with the DS population.

In general, efforts to identify chromosome 21 genes responsible for specific DS phenotypic traits have used two basic approaches. Several groups have studied trisomic mice or partially trisomic mice to examine the effect of increased dosage on the function and regulation of murine genes known to be homologous to those on human chromosome 21 [Reeves et al., 2001; Dierssen et al., 2001]. To date, there is no evidence to link dosage imbalance of a single gene with specific phenotypes using these approaches [Reeves et al., 2001].

The second major approach involves the phenotypic, cytogenetic, and molecular characterization of individuals with partial trisomy of chromosome 21 to create phenotypic maps of chromosome 21 [Korenberg, 1990, 1991; Korenberg et al., 1992, 1994b; Barlow et al., 2001]. The presence or absence of specific abnormalities in each individual with partial trisomy can be correlated with the extent of chromosome 21 present in triplicate. This method can estimate a minimal genomic region responsible for an observed phenotype, although it has a number of limitations. First, there is no direct support for such maps: no single individual has only the smallest region of overlap designated for the trait in triplicate [Reeves et al., 2001]. In addition, analyses are further complicated by the fact that

most individuals with segmental trisomy also have additional chromosomal rearrangements resulting in segmental trisomy or monosomy in other genomic regions. It is, therefore, difficult to conclude that the trisomic region is in fact the sole contributor to the observed phenotype. Lastly, such individuals are rare in the population and thus conclusions are based on a few unique cases. Irrespective, the proposed candidate regions for specific phenotypes based on these patients provide an evidence-based candidate region in which to initiate genetic studies. In the application presented here, we took the candidate approach by prioritizing the study of chromosome 21 genes by their presence in the CHD-critical region: 21q22.2–22.3 between markers D21S55 and COL6A2 [Korenberg et al., 1994a], approximately 8 Mb of candidate sequence.

A complementary approach to mapping genes that may increase susceptibility for variable phenotype is linkage analysis. Given that DS is not passed through pedigrees, classical linkage approaches are not applicable. Instead, Feingold et al. [1995] and Lamb et al. [1996] proposed a method based on a model in which susceptible trisomic genotypes are a result of disomic homozygosity. That is, most susceptible genotypes would be the result of a duplicate copy of the susceptibility allele contributed by the non-disjoining parent; such alleles would be inherited identical by descent. Therefore, this method predicts that individuals with trisomy and the variable phenotype in question would show greater than expected levels of disomic homozygosity in the genomic region containing the susceptibility gene. The limitation of this linkage approach is that it cannot test genetic models that do not lead to increased disomic homozygosity (e.g., those with susceptibility genotypes that act in an additive manner) and a large sample size is needed for those models that lead to only a small increase in disomic homozygosity [Lamb et al., 1996].

An alternative and complementary approach to take in the search for genetic variants would be an association study, where trisomic individuals with a specific defect are identified as "cases" and those without the defect are defined as "controls." To study candidate genes on the trisomic chromosome, modifications to typical genetic association studies must be considered. The greatest concern in the study of a trisomic population is the fact that the assayed alleles are inherited in a non-independent fashion. Due to the nature of the non-disjunction error, trisomic probands inherit either

a representation of each homologous chromosome from the parent in which the error occurred (meiosis I error) or two representations of a single sister chromatid from the same homologue (meiosis II error). This non-independent inheritance of alleles creates problems in many of the techniques utilized for association studies such as allele frequency comparisons, Hardy-Weinberg equilibrium (HWE) analyses, and transmission disequilibrium tests (TDT). In most studies of trisomic populations, the ascertainment of cases and controls is accompanied by the ascertainment of the trisomic individual's parents. The resulting triads (proband, mother and father) provide a wealth of information that can be used in analysis.

There is one additional complication that is unique to a trisomic population; only a small percentage of conceptuses survive to term and, therefore, become eligible for association studies. For example, about 80% of trisomy 21 conceptuses are lost due to significant developmental insults [Freeman et al., 1991]. Therefore, the live birth population from which DS cases and controls may be drawn accounts for only 20% of the trisomy 21 conceptuses. Therefore, we are dealing with a unique subset of the whole DS population that survives to term. We hypothesize that allelic variants in genes related to survival may exist in the live birth DS population. These variants need to be distinguished from those that may lead to susceptibility of the phenotype in question. Alternatively, such genes may be involved in both processes: some alleles may lead to the specific phenotype of interest and others to survival, either related to or unrelated to the phenotype of interest.

In this report, we present a comprehensive strategy to identify genes located on the trisomic chromosome that lead to the increased susceptibility for the variable phenotype and/or the state of survival using an association study. We present the overall strategy for an association study in any trisomic population, including laboratory methods and statistical analyses. We then demonstrate the proposed methodology on a pilot study to identify genes involved in the increased susceptibility for AVSD in a DS population.

STRATEGY

STUDY POPULATION

As in any association study, classification of cases and controls is critical. The variable pheno-

types observed within the DS population show incomplete penetrance and/or variable expressivity. A narrow homogeneous phenotype must be established for the selection of case individuals. For example, almost all major forms of CHD at differing severity levels are observed in the DS population. The best initial strategy may be to select the form of the defect that has the greatest frequency difference between the trisomic and general populations and can be clearly defined. The strictest definition of the control population should include DS individuals with no major birth defect. However, this could be relaxed if the embryonic development of the defect is clearly understood. Once the "case" and "control" definitions are established, trisomic probands are ascertained along with their parents to provide complete triads for analysis.

MOLECULAR METHODS

Most association studies use single nucleotide polymorphisms (SNPs) to characterize the genetic variation found within the gene of interest. Though SNPs are biallelic and, therefore, less informative than microsatellite markers (STRs), they are far more frequent and mutationally more stable. The biallelic nature of SNPs makes them more difficult to accurately genotype in a trisomic population; there are two types of heterozygotes that must be distinguished by dosage (i.e., AAB or ABB). This is not as much of a problem for STRs, as multiallelic patterns are usually clear.

Although there are now several genotyping platforms equipped to quantify the ratio of the number of alleles, we used two SNP genotyping methods we found reliable: pyrosequencingTM (Pyrosequencing, AB; Uppsala, Sweden) and restriction fragment length polymorphism analysis (RFLP) [Wyman and White, 1980]. RFLP analysis was used if a restriction cut site could be identified in the sequence flanking the polymorphic locus. We used WebCutter (<http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html>) to identify possible cut sites. If a restriction cut site was not present, SNPs were genotyped by pyrosequencingTM (Pyrosequencing, AB; Uppsala, Sweden).

For RFLP analysis, 15- μ l reactions included DNA (approximately 35 ng), standard PCR buffer, Mg^{+2} (1.5–2.5 mM), dNTPs (0.2 mM), *FastStart Taq* (1 U, Roche), and primers (0.2 μ M). PCR was preformed in a Perkin Elmer 9700, with an initial denaturation step of 5 min at 95°C, followed by

30 cycles of 95°C for 30 s, Ta for 30 s, 72°C for 30 s, and a final step of 72°C for 5 min. Seven microliters of PCR product was incubated with the appropriate enzyme in a 15- μ l reaction overnight at 37°C. The resultant fragments were separated on a 2.5% agarose gel and visualized by ethidium bromide staining. Genotypes were only scored if products for the parent and trisomic child were separated on the same agarose gel in a side-by-side fashion.

For pyrosequencing analysis, 35- μ l reactions included DNA (approximately 70 ng), standard PCR buffer, Mg^{+2} (1.5–2.5 mM), dNTPs (0.2 mM), *FastStart Taq* (1 U, Roche), and primers (0.2 μ M). PCR was preformed in a Perkin Elmer 9700, with an initial denaturation step of 5 min at 95°C, followed by 45 cycles of 95°C for 30 s, Ta for 30 s and 72°C for 30 s, and a final step of 72°C for 5 min. For all amplicons from both PCR protocols, 5 μ l product was run on a 1.5% agarose gel to confirm amplification. Single-stranded PCR fragments were isolated by using a biotin tag present on a single PCR primer that was conjugated to a streptavidin moiety prior to a denaturation step. Samples were annealed to appropriate sequencing primer and pyrograms (DNA sequence trace) were generated using a MA PSQ96 instrument (Pyrosequencing, AB, Uppsala, Sweden) and the PSQ 96 reagent kit (Pyrosequencing, AB). Genotypes were determined by SNP software V2.0, AQ mode (Pyrosequencing).

STR recombination maps of trisomic offspring based on the genotypes of the mother, father, and

offspring trios can help minimize the possible SNP genotyping error introduced above. STR genotypes and recombination maps are generated as previously outlined [for example, see Feingold et al., 2000]. Figure 1 details an example of the use of STR information to detect SNP genotyping errors. Specifically, examination of alleles from STR markers that flank the candidate gene of interest can establish whether alleles from the parent in whom the error occurred are identical by descent or not. That is, reduction of parental heterozygosity to homozygosity in the child with trisomy would indicate that two alleles were contributed identical by descent from the parent in whom the non-disjunction error occurred. We can use patterns of parent/child STR genotypes to detect genotypic errors in the SNP data, based on the assumption that no double recombinant occurred between STR markers flanking the gene of interest.

STATISTICAL METHODS

There are a number of different analytical methods that can be applied to the data generated by the methods detailed above. We have outlined both preliminary analyses related to the observed pattern of inheritance of the non-disjoined chromosome as well as primary association studies involving the comparison of the case and control trios. In the proposed association study, one can test for HWE in either parents or offspring, compare case and control genotype frequencies

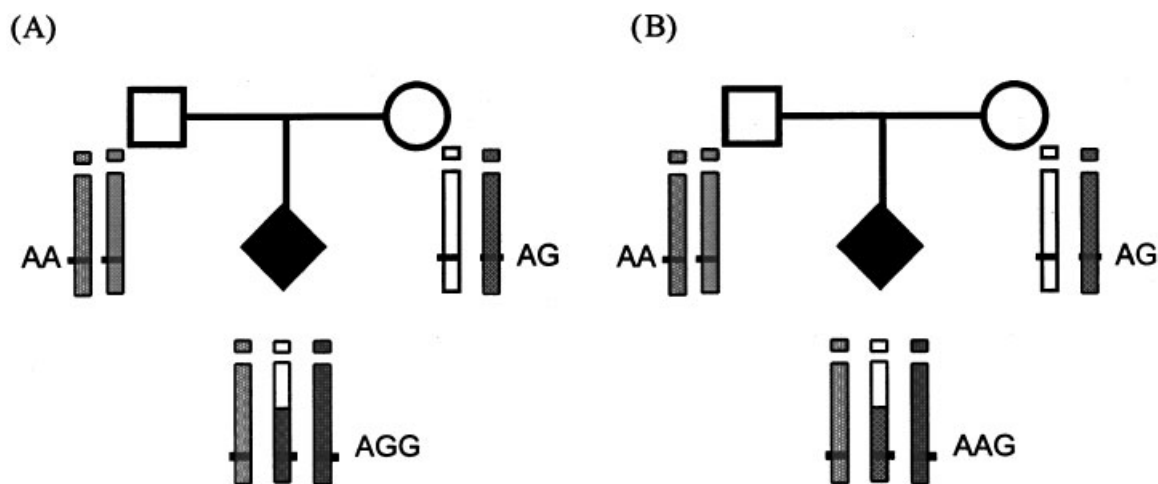


Fig. 1. Utilization of STR data for analysis of genotyping error. Genotyping errors are detected in the proband by analyzing pre-established recombination patterns based on a set of STR markers (represented by the color/pattern of the chromosome). A: Correct SNP genotyping for the proband. B: Potential SNP genotyping error as defined by results that conflict with the STR results.

Proband genotype comparisons:

	HWE		Association tests		
			χ^2	TDT	
	<u>case</u>	<u>control</u>	<u>case/control</u>	<u>case</u>	<u>control</u>
Survival	+	+	n.s.	+	+
Susceptibility for trait	+	n.s.	+	+	n.s.

Parent genotype comparisons:

	HWE		χ^2 association
			<u>case/control</u>
	<u>case</u>	<u>control</u>	
Survival	+	+	n.s.
Susceptibility for trait	+	n.s.	+

Fig. 2. Pattern of results to distinguish candidate gene involvement in survival to term or susceptibility for the trait under study. +: statistically significant result; n.s.: non-significant result.

in parents or offspring, or do a TDT type test [Xu et al., 2004]. Figure 2 summarizes the methods we have considered, along with the expected results at a locus that influences the trait in question. Methods for each of these analyses along with subsequent considerations for a trisomic population will be discussed in the following subsections.

Characterization of the non-disjoined chromosome. The inheritance and recombination pattern of the non-disjoined chromosome can provide important information related to the underlying susceptibility alleles. Preliminary analyses to capture this information include a comparison of the parental origin, meiotic stage of origin, and recombination patterns of the non-disjunction error among those with the defect ("cases") compared with the larger trisomic population or a set of trisomic individuals without the defect ("controls"). The characteristics of the non-disjunction error can be determined by genotyping a panel of highly polymorphic microsatellite (STR) markers along the entire length of chromosome 21 [Lamb et al., 1997].

Differences in the meiotic stage of the error and/or recombination patterns along the non-disjoined chromosome in the cases compared with controls may indicate increased levels of homozygosity in specific chromosome regions. Increased levels of homozygosity can also be examined in the regions flanking candidate genes. Statistical tests to identify differences have been developed previously [Feingold et al., 1995; Lamb et al., 1996].

Parental origin effects would provide initial evidence for the involvement of genes on the trisomic chromosome. For trisomy 21, about 92% of errors are maternally derived. A deviation from the expected ratio of parental errors in the case population may indicate an effect such as imprinting.

HARDY-WEINBERG EQUILIBRIUM

It is important in any association study to test whether the case and control genotypes are in Hardy-Weinberg equilibrium (HWE). This helps detect both genetic effects and genotyping errors. We propose the following method to test HWE in a trisomic population.

The expected probabilities of each of the four genotypes can be calculated as follows, where p and q are the population frequencies of the alleles a and b :

$$\begin{aligned} P(aaa) &= [p^2h + p^3(1-h)], \\ P(aab) &= [pqh + 3p^2q(1-h)], \\ P(abb) &= [pqh + 3pq^2(1-h)], \\ P(bbb) &= [q^2h + q^3(1-h)]. \end{aligned}$$

The parameter h is defined as the probability of disomic homozygosity of the alleles contributed by the non-disjoining parent (NDJ); it has a straightforward relationship to the usual trisomic linkage map parameter. In a trisomy that is a result of a meiosis I non-disjunction, $h=y/2$, where y is the map parameter between the centromere and the marker being tested. In a meiosis II non-disjunction case, $h=1-y$. There is strong evidence for most trisomies that the genetic

map depends on the parent of origin and the stage of origin of the trisomy [Hassold and Hunt, 2001]. For more discussion of the mathematics of genetic maps in trisomy, see, for example, Chakravarti and Slaugenhaupt [1987] or Feingold et al. [2000].

Using the above probabilities, the log-likelihood of a dataset is

$$X_0 \ln[p^2h + p^3(1-h)] + X_1 \ln[pqh + 3p^2q(1-h)] \\ + X_2 \ln[pqh + 3pq^2(1-h)] + X_3 \ln[q^2h + q^3(1-h)],$$

where X_0 , X_1 , X_2 , and X_3 are the observed numbers of offspring with genotypes aaa, aab, abb, and bbb, respectively. It can be shown that the maximum likelihood estimate of p is the usual gene-counting estimate,

$$\hat{p} = \frac{3X_0 + 2X_1 + X_2}{3(X_0 + X_1 + X_2 + X_3)}.$$

The maximum likelihood estimate of h is most easily found by a simple grid search; an analytical solution requires solving a cubic equation. Standard error estimates for \hat{p} and \hat{h} can be obtained, if desired, using asymptotic likelihood theory (i.e., by taking the second derivatives of the log-likelihood), as long as neither parameter estimate is on a boundary.

There are two different ways we can construct a χ^2 test of the null hypothesis of HWE. We can either estimate h as described above, or we can assume it is known from other sources (for example, from a map constructed from a larger dataset). If we estimate h as above, the resulting χ^2 test for HWE has one degree of freedom (four cells, two parameters estimated). If we assume h is known, the test has two degrees of freedom (four cells, one parameter estimated). We prefer the version of the test that estimates h , because if there are genetic effects at the locus, any estimate of h from other sources may be incorrect [Xu et al., 2004].

As usual, HWE should be tested separately in each racial/ethnic group, since the allele frequencies, and thus the expected genotype frequencies, may vary with ethnicity. It is not, however, necessary to test separately by parent of origin or by meiotic stage of origin. The parameter h does depend on parent and stage of origin, and therefore the expected genotype frequencies do as well. However, because the genotype frequencies are *linear* in h , they are correct for a heterogeneous group if h is interpreted as the average value for the group. For example, suppose there are two

subgroups with h -values h_1 and h_2 . Let the frequencies of the two groups be r and $(1-r)$. Then

$$P(\text{aaa}) = r[p^2h_1 + p^3(1-h_1)] + \\ (1-r)[p^2h_2 + p^3(1-h_2)] \\ = p^2[rh_1 + (1-r)h_2] + \\ p^3[r(1-h_1) + (1-r)(1-h_2)] \\ = p^2\bar{h} + p^3(1-\bar{h}).$$

Thus, these two subgroups can be combined in a HWE test as long as the estimated value of h is acknowledged to be an average. One caveat, however, is that if h is estimated elsewhere and treated as "known" in the Hardy-Weinberg test, we must make sure the population from which it is estimated has the same proportions of each parent and stage of origin as the sample in which we are testing HWE.

If deviations from HWE are observed, there are a number of explanations that must be considered. First, genes that may play a role in survival to term and/or in the specific defect being analyzed will lead to altered genotype frequencies in the trisomic population. Examination of HWE in both cases and controls can provide evidence for these hypotheses (Fig. 2). Significant deviation from HWE in both cases and controls would suggest that the gene under study might be involved in survival. Deviation among cases and not among controls would point to the involvement of the gene in the development of the phenotype itself.

Deviation could also be indicative of genotyping errors. As discussed earlier, heterozygote misclassification is a major concern in the analyses of a trisomic population. When significant deviations from HWE are detected, analyses can be repeated collapsing the two heterozygote classes and repeating analyses with three classes. If analysis of the three classes does not produce a significant deviation from HWE, genotyping should be repeated and/or STR flanking markers checked (e.g., Fig. 1) to detect possible errors.

Association study: case-control comparisons. Case-control comparison of proband genotype frequencies can be performed using a χ^2 test for independence (2×4 tables). When χ^2 tests are invalid due to small cell sizes, exact tests can be performed. Analyses must be based on genotypes instead of allele frequencies because inheritance of alleles is non-independent. As in the

Hardy-Weinberg test, the case-control comparison should be done separately by ethnicity, but need not be separated by parent or stage of origin.

It may also be advantageous to do case-control comparisons in the parental populations. The parent in whom the error occurred (NDJ parent) contributes twice as many alleles to the child as the normal disjoining (DJ) parent. Therefore, χ^2 tests for independence could be performed comparing the following pairs: (1) NDJ parent of cases vs. NDJ parent of controls, (2) DJ parent of cases vs. DJ parent of controls, (3) NDJ parent of cases vs. DJ parent of cases, and (4) NDJ parent of controls vs. DJ parent of controls. In these analyses, each pair would reveal different information toward an explanation for a positive association. Comparisons 1 and 2 are direct tests of association, determining if either group of case parents are significantly different from their control counterparts. Comparisons 3 and 4 would reveal if the parents contributing more genetic information (i.e., NDJ parent) are significantly different from their mate. This could either occur in the case population, revealing a positive association with CHD, or in the control population, indicating a potential protective effect. Results from this association test among DS parents and probands in conjunction with the results from HWE and TDT tests (below) provide a pattern that helps distinguish whether the candidate gene contributes to the ability to survive to term or to the confounded effect of survival and susceptibility of the trait under study (Fig. 2).

Transmission Disequilibrium Test. Since trisomic case-control data will typically consist of trios, it is also possible to perform a transmission disequilibrium test (TDT) in cases, controls, or both. This type of test will typically have lower power than a case-control comparison, but will be robust to oddities in the parental population such as population stratification or assortative mating. Xu et al. [2004] describe in detail the trisomic TDT that we have developed. Briefly, it is a likelihood ratio test that looks for non-random segregation of alleles from parents to children while accounting for the unique structure of the trisomic data. A positive test result in *control* trios can be interpreted as evidence of selection effects at the locus being tested (Fig. 2). A positive test result in *case* trios can be interpreted as a confounded effect of the locus on selection and/or on the trait. It is not necessary for the TDT to split the data by ethnicity, parent of origin, or stage of origin.

APPLICATION

We demonstrate the methodologies described above on a pilot study to identify genetic variants that confer increased susceptibility for DS-associated congenital heart defects (CHD).

STUDY POPULATION

In the DS population, all major forms of CHDs are observed in varying degrees. These defects can be thought of as a spectrum of defects, with the end of the spectrum being atrioventricular septal defects (AVSD), often the most severe defect observed. AVSD results from complete failure of septation and communication of all four chambers of the heart. Two forms of AVSD are distinguishable, complete and partial, with the complete form most often associated with DS. We narrowed the phenotypic definition of cases to include only individuals with confirmed cases of complete AVSD. Controls were defined as individuals with trisomy 21 and no major associated defect. Medical records, including echocardiograms, were obtained for both cases and controls to confirm the presence or absence of complete AVSD. Cases and controls were divided into subsets based on broad self-identified racial categories (Caucasian or African-American). The current study population consists of 35 Caucasian and 15 African-American cases. In an effort to increase power, a larger control population was ascertained totaling 50 Caucasian and 22 African-American controls.

Cases and controls for this study were ascertained from two primary sources. The majority were selected from the larger study of live births with free trisomy born and ascertained in the five county metropolitan area of Atlanta, Georgia [Freeman et al., 1998]. Others were ascertained from individuals who attended the DS clinic at Kennedy Krieger in Baltimore, MD. Blood was obtained from mother, father, and proband. Subject recruitment and research protocols were approved by the IRBs of Emory University School of Medicine and the Kennedy Krieger Institution.

CANDIDATE GENE AND MARKER SELECTION

We used the data from Korenberg et al. [1994a] to focus on candidate genes within 21q22.2–22.3. Genes within the CHD-critical region were further narrowed based on their gene expression patterns in the developing fetal heart. At stage E14.5 in the developing mouse embryo, the embryonic heart is

actively developing. Approximately 10% of the genes within the CHD critical region are found to be expressed in developing cardiac tissue at E14.5 [Reymond et al., 2002].

We focused on SH3-binding Glutamine Rich protein (*SH3BGR*). This gene was first identified in a search for genes within the CHD critical region that are expressed in fetal heart tissue [Scartezzini et al., 1997]. Though the function is still unknown, *SH3BGR* is expressed as early as E7 in the mouse heart [Egeo et al., 2000]. From E8.5 to E10.5, when the tubular heart normally undergoes a process of rightward looping, expression of *SH3BGR* is restricted to the developing heart [Egeo et al., 2000]. The process of rightward looping is important in establishing the proper orientation of the heart and thus is important for septation [Srivastava, 2001].

According to Patil et al. [2001], *SH3BGR* is divided into 5 haplotype blocks (B000133, B002020, B002919, B001385, B000139), referred to here as blocks A–E. These haplotype blocks range from just over 54 kb to 300 bp. Blocks A–E contain 5, 3, 2, 4, and 4 unique haplotypes, respectively [Patil et al., 2001]. Our analysis predominately

focused on haplotype blocks B–E, which contain the majority of the *SH3BGR* coding region. We selected ten SNPs for analysis of *SH3BGR* to capture the majority of genetic variation observed in the 20 chromosomes from the global population used in the analysis of Patil et al. [2001]. Markers were termed “gene symbol_a1” (e.g., *SH3BGR_b1*) to reflect the haplotype block and order within a given block in which the SNP may be found. Primers and assay conditions for each SNP are detailed in Table I.

RESULTS

CHARACTERISTICS OF THE NON-DISJOINED CHROMOSOME

Analysis of our samples revealed no statistically significant difference in the parental origin of the non-disjunction error between our case (9/47, 19.2% paternal errors) and control (6/55, 10.9% paternal errors) samples ($P=0.24$). However, if we compare the percentage of paternal errors in the DS case sample with that of our larger population of DS families (55/849, 6.5%), we find a

TABLE I. SNP Genotyping assay descriptions

Polymorphism ^a	dbSNP ^b	PCR primers	Enzyme
A. SNPs genotyped by RFLP analysis			
SH3BGR_b2	rs874221	5'-TGA TAA AGC AA CTG GGA TGA C-3' 5'-CCA AAC CAA GGA AGG AAT AAA-3'	HpaII
SH3BGR_e2	rs2837046	5'-CCA GGA TCA ATC TCA TAT TTT CAG-3' 5'-CCT AGA CAT GAA TGG CAA GTC-3'	HpyCH4IV
SH3BGR_e15	rs1524929	5'-AAT TTT TAG GGA CTG CTC TCA-3' 5'-AAT GCC TCT AAG AAG AAA ACA A-3'	PacI
Polymorphism ^a	dbSNP ^b	PCR primers ^c	Sequence primer
B. SNPs genotypes by Pyrosequencing			
SH3BGR_a1	rs2837010	5'-*GTG AGA AGA TGG GGA AA-3' 5'-AAA TCT TGA TAC TGC CAA AT-3'	5'-CAA CAC ACC CAC CAG-3'
SH3BGR_b1	rs2837041	5'-*GGA GAG GGC TTC TCA ACA TAT C-3' 5'-GGT TGC TCA CTG TAT CAC TCC TT-3'	5'-GTA TTT CTT ATC ACA TAC GG-3'
SH3BGR_b2	rs874221	5'-*TGA TAA AGC AAA CTG GGA TGA C-3' 5'-CCA AA CAA GGA AGG AAT AAA-3'	5'-TGC ACA CTT GTG AGC C-3'
SH3BGR_c1	rs940625	5'-*CAA GGA TTT TGA CTC TTT CTT C-3' 5'-CAT GTA TTA AAC AGT CTT GGA TG-3'	5'-GGA TGC GGT GAT TTC-3'
SH3BGR_d1	rs2837042	5'-CAT AAT GAT ATA TAC CCT TAC A-3' 5'-*AAA TCT GAT AAA ACC TAC CT-3'	5'-AGG TAA GAT CTC GAT TCA A-3'
SH3BGR_d2	rs2837043	5'-*GTA GGG AAG GGA ATG GAG GGT CT-3' 5'-CCT GGA CGG CAA GTG ACT GT-3'	5'-GGC ATG AGG CCA G-3'
SH3BGR_d3	rs2837044	Same amplicon as D2, both SNP analyzed in same fragment	
SH3BGR_e12	rs2837054	5'-TGC TTG CAT TTT GTG GTT CA-3' 5'-BIO/AAA GTA CAG GGA TTA TAG GCG TGA-3'	

^aName given based on position in haplotype blocks established by Patil et al. [2001] (blocks A–E).

^bSNP ID in database dbSNP.

^cDenoted biotin-labeled primer.

statistically significant increase in the frequency of paternal errors in the DS case sample ($P=0.001$).

We also compared the level of homozygosity in the region flanking *SH3BGR* using STR markers. No significant difference between cases and controls was identified; 9/35 (25%) of the informative cases had markers reduced to homozygosity for DS cases compared with 12/64 (19%) for controls ($p=.42$).

TEST OF HARDY WEINBERG EQUILIBRIUM

We tested the null hypothesis of HWE in our cases and controls at each marker using the test described earlier in this study. We report results from the version of the test that estimates h . The “ h known” version of the test gave very similar results, using the value $h=0.25$ estimated from our larger DS population. The P values for the HWE tests are shown for the Caucasian population in Figure 3. Missing symbols for the HWE tests reflect markers where the expected cell sizes were too small to justify the χ^2 approximation (we used a standard of expected values of at least 5.0 in three or more cells). A significant deviation from HWE was observed at marker *SH3BGR_d2* only in the Caucasian control sample (Fig. 3). *SH3BGR_d2* showed an excess of one heterozygote class and a deficit of the other. When the test was repeated with the heterozygotes lumped together, there was no evidence of deviation from HWE. This suggests possible genotyping errors at this marker; however, other explanations that may be considered genotyping errors were not evident in

analyses of the parental samples. In the parental Caucasian sample, a significant departure from HWE was detected only for *SH3BGR_e15*. Because of the small sample size, HWE tests were not valid for the African-American samples.

CASE-CONTROL ASSOCIATION STUDY

The association analysis of *SH3BGR* was performed with nine SNPs. Allele frequency data from our samples differed from the published frequencies for two of the original ten selected SNPs. Marker *SH3BGR_d3*, shown in Patil et al. [2001] to occur with a minor allele frequency of approximately 15%, was monomorphic in our samples and, therefore, uninformative. *SH3BGR_b1* was polymorphic only in our African-American sample. Therefore, only eight SNPs were analyzed in the Caucasian sample; of these, none revealed a significant association (Fig. 3).

In the African-American sample, two SNPs showed a positive association, *SH3BGR_d2* and *SH3BGR_e2* with P values of 0.03 and 0.01, respectively (Fig. 4 and Table II). To examine these associations further, analyses were performed in the parental sample stratified by parental error. Parental comparisons for *SH3BGR_d2* or *SH3BGR_e2* were not statistically different (data not shown).

TRANSMISSION DISEQUILIBRIUM TEST

Most sample sizes in our pilot study were too small to perform the TDT, particularly among the case families. Results for the few TDT analyses

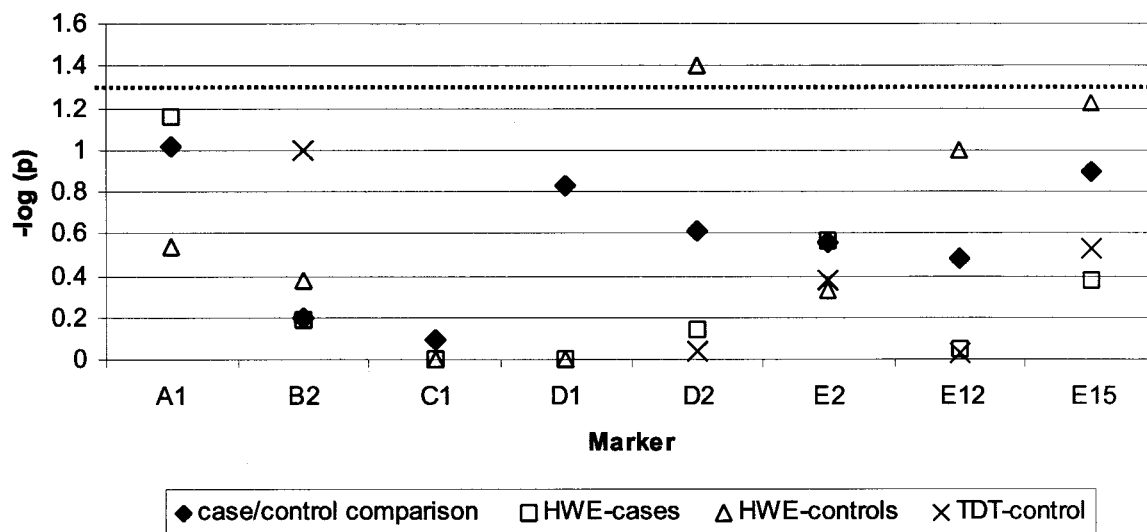


Fig. 3. Association tests for *SH3BGR* and DS-associated AVSD for Caucasian probands. P values are reported as $-\log(p)$; P value for significance ($-\log(0.05)=1.30$) is denoted by the dotted line.

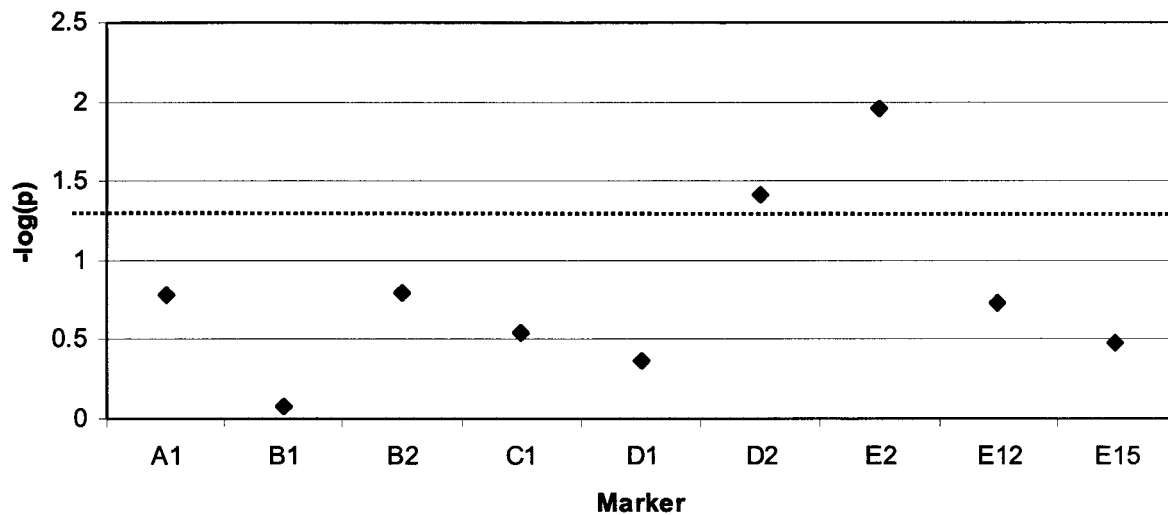


Fig. 4. Case-control association test for SH3BGR and DS associated AVSD for African-American probands (◆). *P* values are reported as $-\log(p)$; *P* value for significance ($-\log(0.05)=1.30$) is denoted by the dotted line.

TABLE II. Genotype frequencies of markers showing significant association

Marker ^a	Genotypes	African-American genotype counts [N (freq.)]		<i>P</i> value ^b	Caucasian genotype counts [N (freq.)]		<i>P</i> value ^b
		Case	Control		Case	Control	
SH3BGR_d2	AAA	12 (0.92)	16 (0.64)	0.03	8 (0.24)	4 (0.10)	0.24
	AAC	0 —	6 (0.24)		12 (0.35)	22 (0.53)	
	ACC	0 —	3 (0.12)		9 (0.26)	4 (0.10)	
	CCC	1 (0.08)	0 —		5 (0.15)	11 (0.27)	
SH3BGR_e2	AAA	5 (0.36)	2 (0.09)	0.01	6 (0.18)	3 (0.06)	0.27
	AAG	4 (0.29)	13 (0.56)		13 (0.38)	17 (0.35)	
	AGG	2 (0.14)	8 (0.35)		8 (0.24)	19 (0.39)	
	GGG	3 (0.21)	0 —		7 (0.20)	10 (0.20)	
SH3BGR_e12	AAA	14 (0.88)	17 (0.074)	0.186	9 (0.31)	7 (0.16)	0.33
	AAG	0 —	4 (0.17)		10 (0.35)	18 (0.40)	
	AGG	1 (0.06)	2 (0.09)		7 (0.24)	10 (0.22)	
	GGG	1 (0.06)	0 —		3 (0.10)	10 (0.22)	

^aName given based on position in haplotype blocks established by Patil et al. [2001] (blocks A–E).

^bPerformed exact test on genotype counts.

among the control families that we were able to perform are shown in Figure 3. Note that the TDTs combine both Caucasian and African-American data. None of the TDTs produced statistically significant results.

DISCUSSION

Little is known about mechanisms leading to the phenotypic variability observed in trisomic populations, particularly DS. To date, no single gene has been associated with any of the variable defects.

Here we present a mapping strategy that complements other genetic approaches to identify genes on the non-disjoined chromosome that lead to susceptibility of trisomy-associated phenotypes. We applied this strategy to a candidate gene, *SH3BGR*, for DS-associated atrioventricular septal defects (AVSD). Genes located on other chromosomes may be involved in trisomy-associated phenotypes when placed in the context of a trisomic background. However, standard linkage and association methods can be used to map such genes.

A trisomic conceptus is the result of a fertilized gamete in which a non-disjunction error has

occurred. The majority of non-disjunction errors occur in oocytes; this is true of almost all chromosomes [Hassold et al., 2001]. However, our case sample revealed a statistically significant increase in the number of paternal non-disjunction errors compared with the overall DS population, 19.1% versus 6.5%. This paternal effect may or may not be related to genes on chromosome 21. To date, there is no evidence that chromosome 21 includes imprinted genes [Rogan et al., 1999; Takai and Jones, 2002]. Although intriguing, this preliminary finding needs to be confirmed in a larger sample of individuals with DS and AVSD.

Case-control association studies were performed based on genotypes instead of allele frequencies to account for the fact that inheritance of alleles in a trisomic population is non-independent. No significant association was detected for any marker in the Caucasian sample. A significant association was detected at two adjacent markers in the African-American sample. One marker in particular, *SH3BGR_d2*, had an increased frequency of heterozygotes among cases compared with controls. As detailed in Figure 2, marginally significant results for these two tests provide preliminary evidence of an association between an *SH3BGR* variant and an increased susceptibility for AVSD in the DS sample and not for survival to term. However, considering the small sizes in both samples, follow-up studies are necessary to confirm such an association.

Most importantly, we have presented a methodology that can be used to detect associations between genes on the non-disjoined chromosome and the variable traits associated with DS. We have explored the effect of survival to term among the sampled DS population and how this might influence results of an association study. The use of the approach presented here in combination with other strategies will provide a comprehensive set of tools to identify genes related to birth defects most commonly observed in individuals with trisomy. Information about the genetic basis of these defects obtained in the trisomic population can then be parleyed into studies toward understanding both normal and abnormal development of the organs in which these defects occur.

ELECTRONIC-DATABASE INFORMATION

dbSNP, <http://www.ncbi.nlm.nih.gov/SNP/Index.html>

Perlegen, <http://www.perlegen.com/haplotypes>
Primer design (pyrosequencing), <http://techsupport.pyrosequencing.com/v2/AssayDesign/index.asp>
Expression profiles, <http://www.tigem.it/ch21exp/>

ACKNOWLEDGMENTS

The authors thank Dr. Ken Dooley, Katherine Allran, and Ignae Thomas for their contributions of case and control ascertainment and Nirupa Masse and Dorothy Pettay for their laboratory assistance.

REFERENCES

- Barlow GM, Chen XN, Shi ZY, Lyons GE, Kurnit DM, Celle L, Spinner NB, Zackai E, Pettenati MJ, Van Riper AJ, Vekemans MJ, Mjaatvedt CH, Korenberg JR. 2001. Down syndrome congenital heart disease: a narrowed region and a candidate gene. *Genet Med* 3:91–101.
- Chakravarti A, Slaugenhaupt SA. 1987. Methods for studying recombination on chromosomes that undergo nondisjunction. *Genomics* 1:35–42.
- Dierssen M, Fillat C, Crnic L, Arbones M, Florez J, Estivill X. 2001. Murine models for Down syndrome. *Physiol Behav* 73:859–871.
- Egeo A, Di Lisil R, Sandri C, Mazzocco M, Lapide M, Schiaffino S, Scartezzini P. 2000. Developmental expression of the *SH3BGR* gene, mapping to the Down syndrome heart critical region. *Mech Dev* 90:313–316.
- Emanuel R. 1970. Genetics and congenital heart disease. *Br Heart J* 32:281–291.
- Epstein CJ. 2001. Down syndrome (Trisomy 21). In: Scriver CR, Beaudet AL, Sly WS, Valle D, editors. *The metabolic and molecular bases of inherited disease*, 8th ed. New York: McGraw-Hill. p 1223–1256.
- Feingold E, Lamb N, Sherman SL. 1995. Methods for genetic linkage analysis using trisomies. *Am J Hum Genet* 56:475–483.
- Feingold E, Brown AS, Sherman SL. 2000. Multipoint estimation of genetic maps for human trisomies with one parent or other partial data. *Am J Hum Genet* 66:958–968.
- Freeman S, Grantham M, Hassold T, Herbert M, Hersey J, Nuccio J, Pettay D, Takaesu N, Phillips C. 1991. Cytogenetic and molecular studies of human spontaneous abortions. *Am J Hum Genet (Suppl)* 49:916A–916A.
- Freeman SB, Taft LF, Dooley KJ, Allran K, Sherman SL, Hassold MJ, Saker DM. 1998. Population-based study of congenital heart defects in Down syndrome. *Am J Med Genet* 80:213–217.
- Fryns JP. 1987. Chromosomal anomalies and autosomal syndromes. *Birth Defects Orig Artic Ser* 23:7–32.
- Hassold T, Hunt P. 2001. To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet* 2:280–291.
- Korenberg JR. 1990. Molecular mapping of the Down syndrome phenotype. *Prog Clin Biol Res* 360:105–115.
- Korenberg JR. 1991. Down syndrome phenotypic mapping. *Prog Clin Biol Res* 373:43–52.
- Korenberg JR, Bradley C, Disteche CM. 1992. Down syndrome: molecular mapping of the congenital heart disease and duodenal stenosis. *Am J Hum Genet* 50:294–302.
- Korenberg JR, Chen X-R, Schipper R, Sun Z, Gonsky R, Gerwehr S, Carpenter N, Daumer C, Dignan P, Disteche C, Graham J,

- Hudgins L, McGillivray B, Miyazaki K, Ogasawara N, Park JPR, Puschel S, Sack G, Say B, Schuffenhauer S, Soukup S, Yamanaka T. 1994a. Down syndrome phenotypes: The consequences of chromosomal imbalance. *Proc Natl Acad Sci USA* 91:4997–5001.
- Korenberg JR, Chen XN, Schipper R, Sun Z, Gonsky R, Gerwehr S, Carpenter N, Daumer C, Dignan P, Disteche C, Graham JM, Hudgins L, McGillivray B, Miyazaki K, Ogasawara N, Park JP, Pagon R, Puschel S, Sack G, Say B, Schoffenhaur S, Soukup S, Yamanaka T. 1994b. Down syndrome phenotypes: the consequences of chromosomal imbalance. *Proc Natl Acad Sci USA* 91:4997–5001.
- Lamb NE, Feingold E, Sherman SL. 1996. Statistical models for trisomic phenotypes. *Am J Hum Gen* 58:201–212.
- Lamb N, Feingold E, Savage A, Avramopoulos D, Freeman S, Gu Y, Hallberg A, Hersey J, Karadima G, Pettay D, Saker D, Shen J, Taft L, Mikkelsen M, Petersen M, Hassold T, Sherman S. 1997. Characterization of susceptible chiasma configurations that increase the risk for maternal nondisjunction of chromosome 21. *Hum Mol Genet* 6:1391–1399.
- Loffredo CA, Hirata J, Wilson PD, Ferencz C, Lurie IW. 2001. Atrioventricular septal defects: possible etiologic differences between complete and partial defects. *Teratology* 63:87–93.
- Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BTN, Noriss MC, Bheehan JB, Shen NSD, Stokowkis RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SPA, Cox DR. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294:1719–1723.
- Reeves RH, Baxter LL, Richtsmeier JT. 2001. Too much of a good thing: mechanisms of gene action in Down syndrome. *Trends Genet* 17:83–88.
- Reymond A, Marigo V, Yaylaoglu MB, Leoni A, Ucla C, Scamuffa N, Caccioppoli C, Dermitzakis ET, Lyle R, Banfi S, Eichele G, Antonarakis SE, Ballabio A. 2002. Human chromosome 21 gene expression atlas in the mouse. *Nature* 420:582–586.
- Rogan PK, Sabol DW, Punnett HH. 1999. Maternal uniparental disomy of chromosome 21 in a normal child. *Am J Med Genet* 83:69–71.
- Scartezzini P, Egeo A, Colella S, Fumagalli P, Arrigo P, Nizetic D, Taramelli R, Rasore-Quartino A. 1997. Cloning a new human gene from chromosome 21q22.3 encoding a glutamic acid-rich protein expressed in heart and skeletal muscle. *Hum Genet* 99:387–392.
- Srivastava D. 2001. Genetic assembly of the heart: implications for congenital heart disease. *Annu Rev Physiol* 63:451–469.
- Takai D, Jones PA. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci USA* 99:3740–3745.
- Wyman AR, White R. 1980. A highly polymorphic locus in human DNA. *Proc Natl Acad Sci USA* 77:6754–6758.
- Xu Z, Kerstann KF, Sherman SL, Chakravarti A, Feingold E. 2004. A trisomic transmission disequilibrium test. *Genet Epidemiol* 26:125–131.
- Zipursky A, Poon A, Doyle J. 1992. Leukemia in Down syndrome: a review. *Pediatr Hemat Oncol* 9:139–149.